Fraud detection with statistics: A comment on *Evidential Value in ANOVA-Regression Results in Scientific Integrity Studies* (Klaassen, 2015).

Hannes Matuschek

Focus Area for Dynamics of Complex Systems and Department of Psychology, University of Potsdam, Karl-Liebknecht-Str. 24, D-14476 Potsdam, Germany, hannes.matuschek@uni-potsdam.de

Abstract

Klaassen in [11] proposed a method for the detection of data manipulation given the means and standard deviations for the cells of a oneway ANOVA design. This comment critically reviews this method. In addition, inspired by this analysis, an alternative approach to test sample correlations over several experiments is derived. The results are in close agreement with the initial analysis reported by an anonymous whistleblower [1]. Importantly, the statistic requires several similar experiments; a test for correlations between 3 sample means based on a single experiment must be considered as unreliable.

1 Introduction

An analysis of means and standard deviations [17], culled from a series of scientific publications, led to a request for retraction of a subset of the papers [16]. The analysis was based on a method reported in Klaassen [11] aimed at detecting a type of data manipulation that causes correlations between condition means of samples that are assumed to be independent. Specifically, given a one-way balanced ANOVA design with 3 conditions, $X_i, i = 1, ..., 3$, the means obtained by averaging over the scores of n different subjects in each condition, are samples of a 3-dimensional normal distribution

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, n^{-1} \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_2 \\ \sigma_1\sigma_2\rho_1 & \sigma_3^2 & \sigma_2\sigma_3\rho_3 \\ \sigma_1\sigma_3\rho_2 & \sigma_2\sigma_3\rho_3 & \sigma_3^2 \end{pmatrix} \right), \quad (1)$$

where μ_i are the unknown *true* expected values and σ_i the unknown sample standard deviations of the scores under the respective conditions and ρ_i their correlations. The ANOVA assumes independence between the samples of the conditions, such that $\rho_i = 0$. Indeed, given only samples of X_i and estimates of σ_i, the sample correlations ρ_i are not directly accessible.

An anonymous whistleblower pointed out [1], that the results in the studies under suspicion (i.e [6], compare Figure 1), show a *super linear* pattern which appears *too good to be true*. Importantly, the authors of the original publications did not necessarily expect such patterns of equidistant means; they expected an ordinal, not a linear relation between the three condition means. Nevertheless, the reanalyses were carried out under the assumption of an expected strict
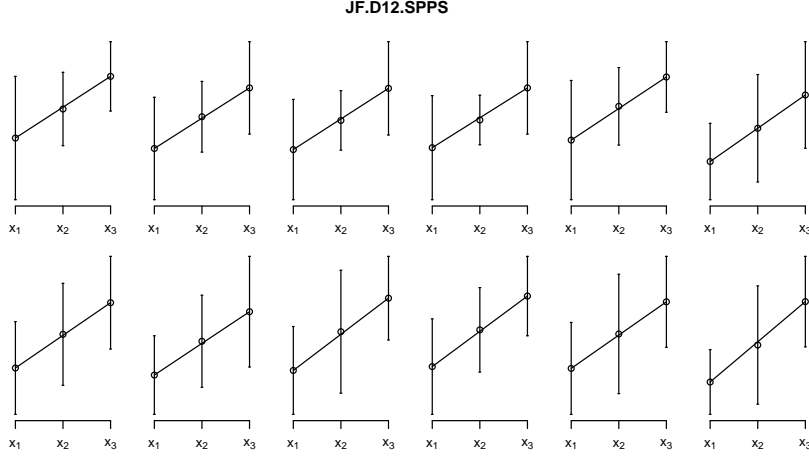
1

JF.D12.SPPS

Fig. 1: Condition means (x_1, x_2 and x_3) and standard deviations for the 12 experiments reported in [6]. The condition means x_1 and x_3 have been connected by a line to visualize the deviance from a perfect linear behavior of the condition means.

linear relation between means. The reason was that this strict assumption is conservative with respect to an inference of data manipulation[1].

Under the assumption of a strictly linear relationship between the group means, $\mu_i = \alpha + \beta \cdot i$, the scores can be described as $X_i = \alpha + \beta \cdot i + \epsilon_i$ which implies that $0 = E[Z] = E[X_1 - 2X_2 + X_3] = \mu_1 - 2\mu_2 + \mu_3$. This linear-combination of sample means X_i yields a new random variable Z with the (univariate) normal distribution $Z \sim \mathcal{N}(0, n^{-1}\sigma_Z^2(\vec{\sigma}, \vec{\rho}))$. Where $\sigma_Z^2(\vec{\sigma}, \vec{\rho}) = \sigma_1^2 + 4\sigma_2^2 + \sigma_3^2 - 4\sigma_1\sigma_2\rho_1 - 4\sigma_2\sigma_3\rho_3 + 2\sigma_1\sigma_3\rho_2$. Note that the random variable Z can be seen as the deviance from the strictly linear behavior $\alpha + \beta\, i$.

Introducing correlations between the samples increases or decreases the variance of Z. Klaassen [11] assumes that a plausible data manipulation (e.g., adjusting the mean of the middle sample towards the mean of the means of the lower and upper samples to achieve significant differences between the groups) leads to a decrease of the variance of Z, $\sigma_Z^2(\cdot, \cdot)$. Such a variance reduction may have gone unnoticed as *humans tend to underestimate variance* in data. As mentioned above, the results under suspicion show a *super linear* behavior and hence a small variance in Z which may not be expected given the group variances σ_i^2 under the assumption of independence.

Consequently, Klaassen [11] used a simple likelihood-ratio test to decide whether there is evidence for data manipulation in terms of a *evidential value* as

$$V = \frac{\max\limits_{\vec{\rho} \in \mathcal{F}} f(z | \sigma_z(\vec{\sigma}, \vec{\rho}))}{f(z | \sigma_z(\vec{\sigma}, \vec{0}))} \ ,$$

comparing the maximum likelihood of all feasible vectors of correlations $\vec{\rho}$ with the likelihood of z under the assumption of $\vec{\rho} = \vec{0}$, where

$$\mathcal{F} = \left\{ \vec{\rho} : \rho_i \in (-1, 1),\ \rho_1^2 + \rho_2^2 + \rho_3^2 - 2\rho_1\rho_2\rho_3 < 1,\ \ \sigma_Z(\vec{\sigma}, \vec{\rho}) \le \sigma_Z(\vec{\sigma}, \vec{0}) \right\} \ ,$$

is the set of feasible correlation vectors, maintaining that the covariance matrix (in eq. 1) remains positive definite and ensures that $\sigma_Z(\vec{\sigma}, \vec{\rho}) \le \sigma_Z(\vec{\sigma}, \vec{0})\, \forall\, \vec{\rho} \in \mathcal{F}$. As the *true* sample

---

[1] It is also not clear how a suitable test could be constructed for the assumption that the means are expected only in a monotonic, not necessarily equidistant order.

standard deviations $\vec{\sigma}$ are unknown, they might be replaced by the reported ones $\vec{s}$, since $\vec{s} \to \vec{\sigma}$ as $n \to \infty$.

2 An asymptotic test statistic

Without any knowledge of the test statistics, i.e. the distribution of V under the null hypothesis H_0 (independent group means), it is not possible to interpret the value V and hence to decide whether a certain value of V does provide evidence for the presence of sample correlations. The estimates of the sample variances $(\vec{s}^2)$ are themselves random variables with some unknown distribution. It is therefore rather unlikely to obtain a closed form expression for the test statistic, even under restrictive assumptions about the distribution of $\vec{s}$.

Nevertheless, as proposed by Klaassen [11], one may assume that asymptotically $\vec{s} \to \vec{\sigma}$ as $n \to \infty$. Then one can assume that the sample variances $\vec{\sigma}$ are fixed and known, allowing for the construction of an upper-bound asymptotic test statistic.

The likelihood to obtain a specific value z, given the sample variances $\vec{\sigma}^2$ and correlations $\vec{\rho}$ is

$$f(Z = z | \sigma_Z(\vec{\sigma}, \vec{\rho})) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma_Z(\vec{\sigma}, \vec{\rho})} \exp\left\{-\frac{n\,z^2}{2\,\sigma_Z^2(\vec{\sigma}, \vec{\rho})}\right\}$$

and therefore

$$V = \max_{\vec{\rho} \in \mathcal{F}} \frac{\sigma_Z(\vec{\sigma}, \vec{0})}{\sigma_Z(\vec{\sigma}, \vec{\rho})} \exp\left\{-\frac{n\,z^2}{2\,\sigma_Z^2(\vec{\sigma}, \vec{\rho})} + \frac{n\,z^2}{2\,\sigma_Z^2(\vec{\sigma}, \vec{0})}\right\}. \tag{2}$$

Now, let $a = \frac{\sigma_Z(\vec{\sigma}, \vec{\rho})}{\sigma_Z(\vec{\sigma}, \vec{0})}$ be the relative standard deviation and $\sigma_0 = \sigma_Z(\vec{\sigma}, \vec{0})$ then

$$V = \max_{a \in \mathcal{A}} a^{-1} \exp\left\{-\frac{n\,z^2}{2\,a^2\sigma_0^2} + \frac{n\,z^2}{2\sigma_0^2}\right\}.$$

The feasible set of all a values $\mathcal{A}$ is implicitly defined by the feasible set of correlations as

$$\mathcal{A} = \left\{\frac{\sigma_z(\vec{\sigma}, \vec{\rho})}{\sigma_z(\vec{\sigma}, \vec{0})} : \vec{\rho} \in \mathcal{F}\right\}.$$

From this it follows immediately that $\mathcal{A} \subseteq (0,1]$ as $\sigma_Z(\vec{\sigma}, \vec{\rho}) \leq \sigma_Z(\vec{\sigma}, \vec{0}) \,\forall\, \vec{\rho} \in \mathcal{F}$.

Under a *worst-case* scenario, one may assume $\mathcal{A} = (0,1]$. This implies that for every $a \in (0,1]$ it is possible to find a feasible correlation vector $\vec{\rho} \in \mathcal{F}$ such that $\sigma_Z(\vec{\sigma}, \vec{\rho}) = a\,\sigma_0$. Please note that this is not ensured in general. The *worst-case* assumption, however, allows one to obtain upper-bounds for the distribution of V under H_0 analytically by relaxing the constraints on a implied by the feasibility constraints on $\vec{\rho}$.

Within this setting one gets

$$V \leq \hat{V} = \max_{a \in (0,1]} a^{-1} \exp\left\{-\frac{n\,z^2}{2a^2\sigma_0^2} + \frac{n\,z^2}{2\sigma_0^2}\right\}.$$

With $\tilde{z} = \frac{\sqrt{n}z}{\sigma_0}$, the normalized z with respect to the expected standard deviation under H_0

$$\hat{V} = \max_{a \in (0,1]} a^{-1} \exp\left\{-\frac{\tilde{z}^2}{2a^2} + \frac{\tilde{z}^2}{2}\right\}.$$

Straightforward computation reveals

$$0 = \partial_a \left(\log\left[a^{-1} \exp\left\{-\frac{\tilde{z}^2}{2a^2} + \frac{\tilde{z}^2}{2}\right\}\right]\right) \quad \Rightarrow \quad a^2 = \tilde{z}^2,$$

and therefore

$$\hat{V} = \begin{cases} 1 & : |\tilde{z}| > 1 \\ |\tilde{z}|^{-1} \exp\left\{\frac{\tilde{z}^2 - 1}{2}\right\} & : \text{else} . \end{cases} \tag{3}$$

Under the *worst-case* scenario, an upper-bound evidential value $\hat{V} \geq V$ can be computed directly without maximizing the likelihood-ratio numerically. This result was also found by Klaassen (compare eq. 18 in [11]).

Knowing that the maximum $\hat{V}$ is achieved at $\tilde{z}^2 = a^2$ and therefore $\frac{nz^2}{\sigma_0^2} = \frac{\sigma_Z^2(\vec{\sigma}, \vec{\rho})}{\sigma_0^2}$, one may conclude that the likelihood-ratio test compares the expected variance σ_0^2 under H_0 with a variance estimated from a single sample. Such a variance estimate is known to be unreliable and therefore the evidential value for a single experiment must be unreliable, too. This issue is discussed in detail in the next section.

3   Testing multiple experiments

Klaassen [11], see also [17] suggested to obtain the evidential value V for an article consisting of more than one experiment as the product of the evidential values V_j of the single experiments in the article. The evidential value V of a publication given N experiments is then

$$V = \prod_{j=1}^{N} V_j = \prod_{j=1}^{N} \max_{\vec{\rho} \in \mathcal{F}_j} \frac{f(z_j | \sigma_Z(\vec{\sigma}_j, \vec{\rho}))}{f(z_j | \sigma_Z(\vec{\sigma}_j, \vec{\rho}))} . \tag{4}$$

Given that $V_j \geq 1$, this immediately implies that the product grows exponentially with the number of experiments even if H_0 is true. Instead of obtaining the evidential value for every single experiment in an article, which (in a worst-case scenario) is based on a variance estimator from a single sample ($\sigma_{Z,j}^2 = n_j z_j^2$), one may try to base that variance estimation on N samples provided by the N experiments in an article. I.e.

$$V = \max_{\vec{\rho} \in \mathcal{F}} \prod_{j=1}^{N} \frac{f(z_j | \sigma_Z(\vec{\sigma}_j, \vec{\rho}))}{f(z_j | \sigma_Z(\vec{\sigma}_j, \vec{\rho}))} , \tag{5}$$

where the feasible set $\mathcal{F} = \bigcap_{j=1}^{N} \mathcal{F}_j$, is just the intersect of all feasible sets $\mathcal{F}_j$ of every experiment.

The idea of this alternative approach is simple: We cannot make a reliable statement about the probability of observing a single *suspiciously* small $\tilde{z}_j$, particularly as $0 = E[Z]$ under H_0. However, observing a *suspiciously* small $\tilde{z}$ repeatedly is unlikely and may indicate sample correlations between groups.

Following the *worst-case* scenario above, the joint evidential value for N experiments is asymptotically

$$\begin{aligned} \hat{V} &= \max_{a \in (0,1]} a^{-N} \exp\left\{ -\sum_{j=0}^{N} \frac{n_j z_j^2}{2\,a^2 \sigma_{0,j}^2} + \sum_{j=0}^{N} \frac{n_j z_j^2}{2\sigma_{0,j}^2} \right\} \\ &= \max_{a \in (0,1]} a^{-N} \exp\left\{ -\sum_{j=0}^{N} \frac{\tilde{z}_j^2}{2\,a^2} + \sum_{j=0}^{N} \frac{\tilde{z}_j^2}{2} \right\} , \end{aligned}$$

where again $\tilde{z}_j = \frac{\sqrt{n_j} z_j}{\sigma_{0,j}}$ and $\sigma_{0,j} = \sigma_Z(\sigma_j, \vec{0})$. A straightforward computation reveals the surprisingly familiar result

$$a^2 = \frac{1}{N} \sum_{j=1}^{N} \tilde{z}^2 \, .$$

This implies that, in a *worst-case* scenario, the joint likelihood-ratio compares a variance estimate based on N samples with the expected one. And finally

$$\hat{V} \;=\; \begin{cases} 1 & : 1 \le \frac{1}{N} \sum_{j=1}^{N} \tilde{z}^2 \\[2ex] \dfrac{\exp\left\{ -\frac{N}{2} + \frac{\sum_{j=0}^{N} \tilde{z}_j^2}{2} \right\}}{\left(\frac{1}{N} \sum_{j=1}^{N} \tilde{z}_j^2 \right)^{\frac{N}{2}}} & : \text{else.} \end{cases}$$

Note that the joint evidential value for N experiments relies on the fact that $\tilde{Z}_j \sim \mathcal{N}(0,1)$ i.i.d. under H_0 and therefore $\sum_{j=1}^{N} \tilde{Z}_j^2 \sim \chi_N^2$. Hence the test statistics for sample correlations between groups can be expressed as a simple chi-squared statistic and one does not need to make the detour of obtaining an approximate distribution of V under H_0.

4   Relation to the ΔF test

The χ^2-test derived in the last section is closely related to the ΔF-test suggested by the whistle-blower [1]. This test was also included in the report for the University of Amsterdam [17].

Under H_0 and the assumption of a linear trend, the p-values of the ΔF-test for a single experiment within an article are distributed uniformly in $[0, 1]$. Using Fisher's method, it is then possible to obtain a p-value for an article comprising several experiments. The major difference between these two methods is that the ΔF-test first determines a p-value for every study and tests whether the resulting p-values p_j are *to good to be true* while the chi-square test introduced here assesses this value directly by inspecting whether the relative deviations form perfect linearity $\tilde{z}_j^2$ are *to good to be true*. Therefore, unsurprisingly, the two methods yield very similar results (see Table 1).

Article	χ^2-test	ΔF-tests	Classification
JF09.JEPG [3]	8.06e-07	2.30e-07	strong
JF11.JEPG [5]	8.73e-07	3.53e-07	strong
JF.D12.SPPS [6]	7.14e-09	1.82e-08	strong
L.JF09.JPSP [13]	6.44e-4	8.46e-5	strong
L.JF09.JPSP*	0.03	0.02	–
JF.LS09.JEPG [4]	0.25	0.11	strong
JF.LK08.JPSP [7]	0.81	0.66	inconclusive
D.JF.L09.JESP [2]	0.93	0.52	inconclusive
Reference [8, 9, 10, 12, 15, 18, 20, 22, 21]	0.11	0.14	–

Tab. 1: Comparison of p-values obtained with the direct χ^2 and ΔF tests for studies classified as providing strong or inconclusive statistical evidence for low veracity by Peeters et al. [17]. The first three studies listed in the table were reported by the whistleblower [1]. Note the divergence for JF.LS09.JEPG between the present analysis and [17]. Only those studies from [17] were considered here which provide at least 8 experiments.
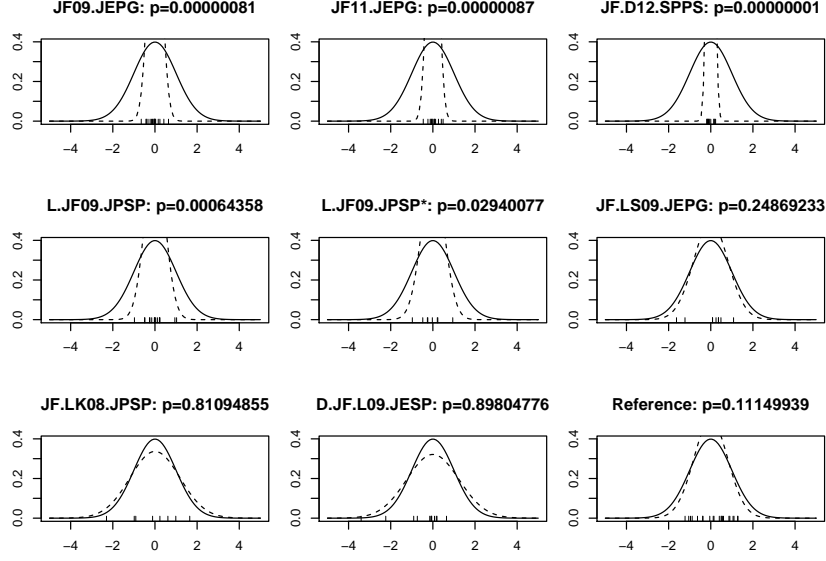
Fig. 2: The distribution of $\tilde{z}_j$ (short dashes at the bottom of each panel) for each experiment from the articles listed in Table 1. The solid line shows the expected distribution of $\tilde{Z}_j$ under H_0 while the dashed line shows the normal distribution with 0-mean and the variance estimated from the samples $\tilde{z}_j$.

Both methods, the χ^2 and ΔF tests, are *conservative* compared to the V-value approach by Klaassen [11]. For example, the article JF.LS09.JEPG in Table 1 was classified with *strong statistical evidence for low veracity* [17] (compare also Figure 3). In contrasts, the χ^2 and ΔF methods, yield p-values of ≈ 0.25 and ≈ 0.11, respectively, suggesting that there is no evidence of sample correlations between groups. The three methods agree for the studies JF.LK08.JPSP and D.JF.L09.JESP which were classified with *inconclusive statistical evidence for low veracity*. The three methods also agree on classifying the three articles reported by the whistleblower [1] with *strong statistical evidence for low veracity.*

Depending on the chosen level of significance, the article L.JF09.JPSP could be classified as *strong* or *inconclusive.* This article contains conditions for which the authors did not expected a specific rank ordering of the condition means. Peeters et al. [17] included these *control conditions* but reordered them according to increasing group means, yielding a p-value for the χ^2-test of about 0.0006 (L.JS09.JPSP in Table 1). Although the assumption of equidistant group means, i.e. $0 = \mu_1 - 2\mu_2 + \mu_3$, contains the assumption of equal group-means, i.e. $\mu_1 = \mu_2 = \mu_3$ as a special case, the actual test-result depends on the ordering of the conditions. Keeping the order of conditions as reported in [13] yields a p-value of about 0.015 and excluding them results in a p-value of about 0.03, shown as L.JF09.JPSP* in Table 1.

The discrepancy between the χ^2 or ΔF methods and the V-value method for the JF.LS09.JEPG article [4] is due to the tendency of the V-value method to indicate *strong evidence* if a single experiment out of a series of experiments has a very small $\tilde{z}$-value. In contrast to the V-value method, the χ^2 and the augmented V-method (see Section 3) take all experiments of an article into account by assuming the same correlation structure for all experiments.

For the particular article [4], the V-value approach reported *strong evidence for low veracity* because the last two experiments (compare Figure 3) exhibit the *super linear* pattern associated
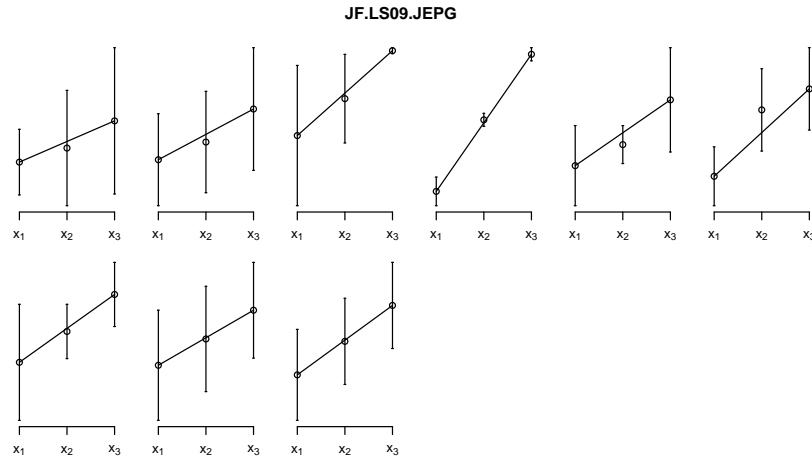
JF.LS09.JEPG

Fig. 3: Condition means and stdandard deviations for 9 experiments from [4].

with sample correlations. The χ^2 and ΔF method, however, do not indicate significant sample correlations as the deviance of remaining experiments fit well into the expected distribution under H_0, especially the results in panels 5 & 6 in Figure 3.

Klaassen [11] intended the V-value to be sensitive for single experiments. The argument is that *bad science cannot be compensated by very good science* [11]. Finding a small value for $\tilde{z}_j$ in a series of experiments, however, is quiet probable[2] even under H_0. Hence one could argue that a single *suspiciously small* $\tilde{z}_j$ can not be interpreted as strong evidence for sample correlations.

5 Discussion

There is no doubt that, in principle, statistics can be used to detect sample correlations that are due to data manipulation. The approach proposed in [11], however, is not without problems.

A first problem is the missing test statistics for the evidential value V. Although an upper-bound asymptotic test statistics for the V-value of a single experiment can be obtained (see Section 2 above and [11]), the reliability of the V value for a small n remains unknown (as well as how large a large n must be to be considered *large*).

A second problem is the critical value of $V^* = 6$ chosen by the authors, which implies (asymptotically) $p \approx 0.08$. Arguably, this is a rather high probability of falsely accusing a colleague of data manipulation.

A third problem is the assumption that the product of the evidence provided by every single experiment in an article can serve as a metric of evidence for data manipulation in this article. As mentioned above as well as in the comments to the article at pubpeer.com [19] and in a response by Denzler and Liberman [14], this assumption implies that the evidence for data manipulation grows exponentially with the number of experiments even under H_0. The probability of $V \geq 2$ for a single experiment is about $p \approx 0.25$. Thus, about every 4th *good* experiment will double the evidence for data manipulation.

The fourth problem, finally, is a general concern. The analysis assumes a specific type of data manipulation. If this is true, the manipulation will induce correlations between condition means. Moreover, under the second assumption that $0 = X_1 - 2X_2 + X_3$ this correlation can

---

[2] I.e. for 10 experiments ($N = 10$) $p \approx 0.4$ for $\alpha = 0.05$ and $p \approx 0.1$ for $\alpha = 0.01$

be detected. Importantly, however, the reverse is not true: The detection of such correlations in the data does not necessarily imply that data were manipulated. For that reason, Peeters et al. carefully avoided in [17] to claim that their findings prove that data were manipulated. Instead the results are interpreted as *evidence for low data veracity*, which is justified. In [11], however, Klaassen claims that its method provides evidence for manipulation. Although the origin of sample correlations cannot be determined with statistics, their presence certainly violates an ANOVA assumption. This may result in an increased type-I error rate. Therefore, the effects reported in the articles providing strong or possibly even inconclusive evidence for sample correlations (e.g [3, 5, 6, 13]) may be less significant than suggested by their ANOVAs.

In this comment, specifically in Section 3, the concept of the single-experiment evidential value was extended to multiple experiments. Moreover, a much simpler chi-squared test was provided to test the presence of correlations in the data that is similar to the test proposed in [1] and yielded very similar probabilities for the presence of sample correlations. Thus, the V-value approach can serve as a test for sample correlations, if it is applied across several identical or at least similar experiments. In this case one is also able to decide whether the variability in the results is suspiciously small or not. However, estimating σ_Z on the basis of a single experiment will certainly not reveal a reliable result.

References

[1] Anonymous. Suspicion of Scientific misconduct by Dr. Jens Förster, 2012.

[2] Markus Denzler, Jens Förster, and Nira Liberman. How goal-fulfillment decreases aggression. *Journal of Experimental Social Psychology*, 45(1):90–100, 2009.

[3] J. Förster. Relations between perceptual and conceptual scope: How global versus local processing fits a focus on similarity versus dissimilarity. *Journal of Experimental Psychology: General*, 138:88–111, 2009.

[4] J. Förster, N. Liberman, and O Shapira. Preparing for novel versus familiar events: Shifts in global and local processing. *Journal of Experimental Psychology: General*, 138:383–399, 2009.

[5] Jens Förster. Local and global cross-modal influences between vision and hearing, tasting, smelling, or touching. *Journal of Experimental Psychology: General*, 140:364–389, 2011.

[6] Jens Förster and Markus Denzler. Retracted: Sense Creative! The Impact of Global and Local Vision, Hearing, Touching, Tasting and Smelling on Creative and Analytic Thought. *Social Psychological and Personality Science*, 3(1):118–118, 2014.

[7] Jens Förster, Nira Liberman, and Stefanie Kuschel. The effect of global versus local processing styles on assimilation versus contrast in social judgment. *Journal of personality and social psychology*, 94(4):579–599, 2008.

[8] H. Hagtvedt and V. M. Patrick. Turning Art Into Mere Illustration: Concretizing Art Renders Its Influence Context Dependent. *Personality and Social Psychology Bulletin*, 37(12):1624–1632, 2011.

[9] Catherine Hunt and Marie Carroll. Verbal overshadowing effect: How temporal perspective may exacerbate or alleviate the processing shift. *Applied Cognitive Psychology*, 22(1):85–93, 2008.

[10] A. B. Kanten. The effect of construal level on predictions of task duration. *Journal of Experimental Social Psychology*, 47(6):1037–1047, 2011.

[11] Chris A. J. Klaassen. Evidential Value in ANOVA-Regression Results in Scientific Integrity Studies. pages 1–12, 2015. arXiv:1405.4540 [stat.me].

[12] Davy Lerouge. Evaluating the Benefits of Distraction on Product Evaluations: The Mind-Set Effect. *Journal of Consumer Research*, 36:367–379, 2009.

[13] N. Liberman and J. Förster. Distancing from experienced self: How global-versus-local perception affects estimation of psychological distance. *Journal of Personality and Social Psychology*, 97:203–216, 2009.

[14] Nira Liberman and Markus Denzler. Response to a Report Published by the University of Amsterdam The, 2015.

[15] Selin A. Malkoc, Gal Zauberman, and James R. Bettman. Unstuck from the concrete: Carryover effects of abstract mindsets in intertemporal preferences. *Organizational Behavior and Human Decision Processes*, 113(2):112–126, 2010.

[16] University of Amsterdam. Articles jens förster investigated, 2015.

[17] Carel F. W. Peeters, Chris A. J. Klaassen, and Mark A. van de Wiel. Evaluating the Scientific Veracity of Publications by Dr. Jens Förster, 2015.

[18] Evan Polman and Kyle J Emich. Decisions for others are more creative than decisions for the self. *Personality and social psychology bulletin*, 37(4):492–501, 2011.

[19] Pubpeer.com. Evidential Value in ANOVA-Regression Results in Scientific Integrity Studies, 2015.

[20] Laurens Rook and Daan van Knippenberg. Creativity and Imitation: Effects of Regulatory Focus and Creative Exemplar Quality. *Creativity Research Journal*, 23:346–356, 2011.

[21] Pamela K Smith and Yaacov Trope. You focus on the forest when you're in charge of the trees: power priming and abstract information processing. *Journal of personality and social psychology*, 90(4):578–596, 2006.

[22] Pamela K. Smith, D. H J Wigboldus, and Ap Dijksterhuis. Abstract thinking increases one's sense of power. *Journal of Experimental Social Psychology*, 44(2):378–385, 2008.